

研究ノート

## 留学生の日本語能力とライティング能力に関する基礎的研究

—プレースメントテストと自動評価サイトを使って—

Basic Study on Japanese Language Ability and Writing Skills for International Students:  
Using Placement Test and Automated Evaluation

坪根由香里\*

TSUBONE Yukari

This paper analyzes the relationship between Japanese language ability, the automated evaluation results of writing, and text data such as the total number of characters and sentence length, using the results of an international student's Japanese placement test at a university and automated evaluation sites. As a result, there was a correlation between Japanese language ability and automated evaluation results, students with higher Japanese proficiency increased the total length of writing, the length of sentences, and the number of morphemes and even students with high Japanese ability could not make long and complex sentences like native speakers of Japanese. Teaching approaches for writing based on the results were considered.

キーワード: ライティング評価 (writing evaluation)、日本語能力 (Japanese language ability)、ライティング能力 (writing skills)、プレースメントテスト (placement test)、自動評価 (automated evaluation)

## 1. はじめに

留学生が大学生活を送る上で、ライティング能力は不可欠なものである。第二言語（以下、L2）ライティング能力は、文章能力（内容や構成に関わる能力）と言語能力に分けられる。言語能力について考えると、その要素には、文法・語彙・表記等の正確さに加え、まとまった分量のものを書く能力、ある程度の長さの文を作る能力、語彙の豊富さ、漢字の適切な使用等が挙げられる。作文を書く際に必要なそれらの要素（以下、テキストデータ）は、プレースメントテスト等で実施される筆記試験で測る言語能力（日本語能力）とどのような関係があるのだろうか。また、言語能力（日本語能力）と文章能力は関係があるのだろうか。

近年、自動評価（機械評価）の開発が進んでおり、上記のようなテキストデータについても、一部、自動評価サイトで情報が提供されるようになっている。自動評価結果の精度に関しては完全ではない部分もあるとされているが、特に大人数のライティング能力を測る場合や、

作文クラスにおける学生へのフィードバックには、それらの情報を効果的に使うことも必要であろう。そのためには自動評価サイトで出される結果についても検証する必要がある。そこで、以下の4つを研究課題とする。

課題 1: 日本語能力と作文の自動評価結果は関係があるのか。日本語能力と文章能力には関係があるのか。

課題 2: 各自動評価は関係があるのか。

課題 3: 作文の総合的評価とテキストデータにはどのような関係があるのか。

課題 4: 日本語能力とテキストデータにはどのような関係があるのか。

## 2. 先行研究

日本語学習者のライティングに関する研究は、主に文章能力に関するものと、言語能力に関するものに分けられる。

文章能力に関するものには、田中・坪根 (2011)、坪根・田中 (2015)、田中他 (2017) 等がある。田中・坪根 (2011)

\*大阪観光大学観光学部/日本語教育

は、第二言語としての日本語小論文における Good writing は、日本語教師にどのように捉えられているのかを調査したもので、「課題の達成」「主張の明確さ」「内容のオリジナリティ」「客観的で広い視野からのサポート」「構成」「談話展開のテクニック」「表現力の豊かさ」等が Good writing の順位決定要素となっていること等を明らかにしている。また、坪根・田中 (2015) は、第二言語としての日本語小論文の「いい内容」「いい構成」の要因を示している。田中他 (2017) では、「多言語母語の日本語学習者横断コーパス」(I-JAS) のエッセイを使い、1~6 レベルの総合的評価、およびマルチプルトレイト (以下、MT) 評価 (「目的・内容」「構成・結束性」「日本語」の 3 つのトレイト別評価) を行っている。そのうち、「目的・内容」については、総合的評価と差があるものについて質的に分析し、「構成・結束性」については、主にメタ言語や構成の型について質的に分析している。

言語能力に関するものとしては、長谷川 (2004) が、大学のプレースメントテストで実施された作文の教師による評価と作文の形態的特徴の関係について調査している。その結果、作文評価は延べ語数・異なり語数と高い相関があること、SPOT<sup>2)</sup>や音声や文法などのマークシートテストとの間に高い相関が見られたことを示した。田中他 (2017) は総合的評価によって上位群・中位群・下位群に分け、日本語の言語特徴量を分析している。その結果、上位群は作文全体が長く、一文も長くなっていること、上の群ほど異なり形態素数も総形態素数も多くなること等を明らかにしている。

日本語ライティングの自動評価については、代表的なものとして、Goodwriting Rater、jReadability、jWriter がある。Goodwriting Rater は、「比較と意見 (論証)」のライティング (作文) を自動評価 (機械評価) し、その評価結果 (ホリスティック評価 (総合的評価) と MT 評価 (「目的・内容」「構成・結束性」「日本語」)) を示すとともに、テキストに関する情報を提供し、メタ言語をハイライトして示すものである。評価結果とともに、その評価になる確率も「確信度」として提示している。jReadability は、テキストの難易度を 6 段階で示し、語彙レベル構成、品詞構成、語種構成、文字種構成などの情報を表示する。また、初級前半~上級後半の 6 段階に分けた「語彙レベル構成」も提示している。李 (2016) よると、このシス

テムは「高精度で文章の難易度を予測しうる」という。また、jWriter は、作文テキストを入力すると、推測される作文力のレベルが 5 段階で判定され、語の多様性、漢語力、長文作成力、難解語についての評価も示される。

本研究では、大学における留学生の日本語プレースメントテストの結果と、Goodwriting Rater、jReadability、jWriter を利用し、日本語能力、作文の自動評価結果、テキストデータの間関係を分析する。

### 3. 調査の概要

#### (1) 調査協力者

調査は、関西にある A 大学に 2018 年 4 月に入学した外国人留学生 140 名を対象とした。出身国・地域の内訳は、中国 97 名 (うち香港 1 名)、ベトナム 31 名、韓国 5 名、台湾 4 名、ネパール 2 名、インドネシア 1 名である。

#### (2) 調査の方法

2018 年 4 月 2 日に上記留学生に対して実施されたプレースメントテストを分析対象とする。プレースメントテストでは、文法・読解を中心とした日本語テスト (以下、日本語テスト) と作文テストの 2 種類のテストが実施された。作文のテーマは以下の通りである。時間は 60 分で、辞書の使用は認めていない。

テーマ: あなたは「<sup>がいしょくほ</sup>外食派」? それとも「<sup>じすいほ</sup>自炊派」?  
 「<sup>がいしょく</sup>外食 (外で食事をする事)」と「<sup>じすい</sup>自炊 (家で自分で食事を作って食べる事)」、それぞれのプラス面 (いい点) とマイナス面 (悪い点) を挙げて比較し、「<sup>めん</sup>食生活」についてのあなたの<sup>いけん</sup>意見を 600 字~800 字で書いてください。

分析手順は以下の通りである。

- ① 作文を Goodwriting Rater、jReadability、jWriter の各自動評価サイトで評価し、同時に提示されるテキストデータ (総文字数、総段落数、総文数、異なり形態素数、文あたりの平均文字数等) を収集する。
- ② 日本語テストの得点と自動評価サイトの各評価、テキストデータとの相関を調べる。
- ③ Goodwriting Rater の評価 (ホリスティック評価) と jReadability、jWriter の評価、テキストデータとの相関を調べる。
- ④ 日本語テストによって上位群・中位群・下位群に分

け、テキストデータにはどのような差があるのかを調べる。

⑤そこから作文指導に必要なものについて考察する。

#### 4. 結果と考察

##### (1) 日本語能力と作文の自動評価結果の関係

日本語テストの得点と自動評価サイトの各評価<sup>3)</sup>の相関関係を、スピアマンの順位相関係数を用いて分析した(表-1)。

表-1の「ホリスティック評価」と「MT評価」は Goodwriting Rater によるものである。「jReadability 文章難易度」および「リーダビリティスコア」は jReadability によるもので、リーダビリティスコアは数値が大きい場合、相対的に易しいテキストであるとされている。「jWriter」は jWriter による評価である。以下、\* p<.05、\*\* p<.01、\*\*\* p<.005、\*\*\*\* p<.001 として示す。

表-1 日本語テストと各自動評価の相関

ホリスティック評価	MT評価 目的・内容	MT評価 構成・ 結束性	MT評価 日本語
.552**	.251**	.558**	.324**

  

jReadability 文章難易度	リーダビリ ティスコア	jWriter
.322**	-.505**	.543**

日本語テストは全ての自動評価の結果と相関が見られ、日本語テストの得点が高ければ、自動評価による評定値も高くなり、文章の難易度も上がるということがわかった。作文を評価するこれらの自動評価は日本語能力の影響が強いということがうかがえる。また、この自動評価では、「目的・内容」や「構成・結束性」といった文章能力も日本語テストとの相関が認められた。

##### (2) 各自動評価の相関

次に、各自動評価サイトの結果が互いに関係があるのを見るため、スピアマン順位相関係数を用いて、Goodwriting Rater のホリスティック評価と jReadability、jWriter の評価の間の相関を調べた(表-2)。

表-2 Goodwriting Rater (ホリスティック評価)、jReadability、jWriter の相関

	Goodwriting Rater	jReadability		jWriter
		文章難易度	リーダビリ ティスコア	
Goodwriting Rater	—			
jReadability	文章難易度	.522**	—	
	リーダビリ ティスコア	-.603**	※	—
jWriter	.552**	.572**	-.840**	—

※文章難易度とリーダビリティスコアはともに jReadability によるものであるため、相関は出していない。

以上の結果からは、Goodwriting Rater のホリスティック評価、jReadability、jWriter による評価は互いに相関があることがわかった。

##### (3) 作文の総合的評価とテキストデータの関係

作文の様々な数的データが作文の総合的評価に影響を与えているかどうかを調べるために、ここでは自動評価の中から Goodwriting Rater のホリスティック評価を用いて、各テキストデータとの相関係数を求めた(表-3)。以下、「総文字数」「総文数」「総段落数」「漢字率」「一文の平均文字数」は Goodwriting Rater、「一文の平均語数」「総形態素数(延べ)」「総形態素数(異なり)」は jReadability のデータを用いている。

表-3 Goodwriting Rater (ホリスティック評価)と各テキストデータの相関

総文字数	総文数	総段落数	漢字率
.372**	-.271**	-.128	.040

  

一文の 平均文字数	一文の 平均語数	総形態素数 (延べ)	総形態素数 (異なり)
.410**	.404**	.330**	.428**

Goodwriting Rater のホリスティック評価は、総文字数、総文数、一文の平均文字数、一文の平均語数、総形態素数(延べ・異なり)と相関があった。つまり、作文の総合的評価が高いものは、作文全体の長さや一文の長さが長くなる一方で文数は減り、形態素の数が多くなると言える。総段落数、漢字率との相関は見られなかった。

##### (4) 日本語能力とテキストデータの関係：日本語テスト

### の上位群・中位群・下位群の比較

日本語能力と作文の数的データとの関係を調べるために、まず、日本語テストとテキストデータの相関を調べた (表-4)。

表-4 日本語テストと各テキストデータの相関

総文字数	総文数	総段落数	漢字率
.249**	-.327**	-0.072	-0.133
一文の平均文字数	一文の平均語数	総形態素数 (延べ)	総形態素数 (異なり)
.425**	.425**	.234**	.465**

日本語テストの得点も、(3)の結果同様、総文字数、総文数、一文の平均文字数、一文の平均語数、総形態素数 (延べ・異なり) の間に相関が見られた。つまり、日本語能力が高ければ、作文全体の長さや一文の長さが長くなる一方で文数は減ること、形態素の数が増えることがわかった。また、漢字率と日本語テストの相関は認められなかった。

次に、日本語テスト (50 点満点) の得点から、上位群 (48 名)・中位群 (48 名)・下位群 (44 名) に分け、各テキストデータについて差があるのかを調べた。まず、3つの群が異なるグループと言えるかどうかを調べるために、日本語テストについて一元配置分散分析 (ANOVA4 on the Web 使用) を行った。主効果 ( $F(2,137)=414.850$ ,  $p=.000$ \*\*\*\*) が有意であったため、ライアン法による多重比較を行ったところ、上位群と中位群、上位群と下位群、中位群と下位群すべての間に有意差があった。よって、3つの群は日本語テスト得点において異なるグループであると言える。各群の平均値と標準偏差 (SD) を表-5 に示す。

表-5 上位群・中位群・下位群の日本語テストの平均値・標準偏差

	平均 (SD)
上位群 (n=48)	35.02 (4.26)
中位群 (n=48)	22.67 (2.38)
下位群 (n=44)	13.66 (3.73)
全体 (n=140)	24.07 (9.40)

次に、各テキストデータについて、上位群・中位群・

下位群の間に差があるのかを、一元配置分散分析によって調べた。表-6 は3つの群の各項目の平均値と標準偏差 (SD) を示したものである。

主効果はすべての項目において優位であった (総文字数:  $F(2,137)=5.148$ ,  $p=.007$ \*\*、総文数:  $F(2,137)=7.810$ ,  $p=.0006$ \*\*\*\*、一文の平均文字数:  $F(2,137)=7.032$ ,  $p=.0012$ \*\*\*、一文の平均語数:  $F(2,137)=8.424$ ,  $p=.0004$ \*\*\*\*、総形態素数 (延べ):  $F(2,137)=5.154$ ,  $p=.0069$ \*\*、総形態素数 (異なり):  $F(2,137)=18.784$ ,  $p=.0000$ \*\*\*\*)。

表-6 上位群・中位群・下位群の各テキストデータの平均値・標準偏差

	総文字数 平均 (SD)	総文数 平均 (SD)	一文の平均文字数 平均 (SD)
上位群	628.38 (72.41)	19.50 (4.58)	33.64 (8.19)
中位群	598.31 (73.98)	20.92 (5.50)	30.32 (9.01)
下位群	579.84 (71.48)	24.41 (7.74)	26.39 (10.28)
全体	602.81 (75.34)	21.53 (6.37)	30.22 (9.63)
	一文の平均語数 平均 (SD)	総形態素数 (延べ) 平均 (SD)	総形態素数 (異なり) 平均 (SD)
上位群	21.10 (4.97)	394.23 (50.26)	138.15 (17.74)
中位群	18.99 (5.45)	375.23 (50.24)	124.04 (16.47)
下位群	16.38 (5.93)	361.25 (46.17)	115.30 (19.52)
全体	18.89 (5.78)	377.35 (50.81)	126.13 (20.21)

それぞれライアン法 (5%水準) による多重比較を行ったところ、以下のような結果になった。

総文字数: 上位群 > 下位群

総文数: 上位群 = 中位群 < 下位群

一文の平均文字数: 上位群 > 下位群

一文の平均語数: 上位群 = 中位群 > 下位群

総形態素数 (延べ): 上位群 > 下位群

総形態素数 (異なり): 上位群 > 中位群 > 下位群

この結果からは、上位群は下位群より作文全体 (総文字数) が長く、一文の平均文字数および一文の平均語数が多いことから、一文の長さも長いと言える。総文数が上位群・中位群より下位群が多いことと合わせて考えると、下位群は短文の羅列で、日本語力が上がるほど長い文が増えるために総文数が少なくなると考えられる。また、総形態素数 (異なり) が下位群、中位群、上位群の順に多くなることから、日本語力が上がるにつれて語彙も豊富になると言える。

しかし、上位群と中位群の間に差があったのは、総形態素数 (異なり) のみであった。総文字数に関しては、

文字数を 600 字～800 字としたため、天井効果があった可能性がある。一文の平均文字数については、日本語母語話者による 800 字の小論文について調べた原田(2019)によると、日本語母語話者の書く文は一文あたり「71 文字から 120 文字のものが多い」という。適切な文の長さについては諸説あるが、本調査における上位群平均の 33.64 字は日本語母語話者と比べると短いと言えるだろう。このことから、上位群であっても長く複雑な文の生成ができるまでには至っていないと推察される。

上記データの他に、jReadability では初級前半～上級後半の 6 段階に分けた「語彙レベル構成」も示している。表 7 に各段階の語彙の割合の平均値と標準偏差 (SD) を示した。なお、上級前半、上級後半は数が少なくて割合が表示されないものもあったため、分析から除外する。

表 7 上位群・中位群・下位群のレベル別語彙割合の平均値・標準偏差

	初級前半	初級後半	中級前半	中級後半
	平均(SD)	平均(SD)	平均(SD)	平均(SD)
上位群	32.71(5.25)	33.94(5.01)	20.77(5.23)	10.75(3.81)
中位群	34.13(4.84)	37.13(6.18)	17.10(4.37)	9.04(3.27)
下位群	38.55(6.49)	35.80(6.06)	14.00(4.55)	7.80(3.16)
全体	35.03(6.06)	35.61(5.94)	17.39(5.48)	9.30(3.56)

各レベルの語彙の割合が 3 つの群の間で異なるのかを一元配置分散分析で調べたところ、初級前半、初級後半、中級前半、中級後半において主効果が有意であった (初級前半:  $F(2,137)=13.410$ ,  $p=.0000^{****}$ 、初級後半:  $F(2,137)=3.594$ ,  $p=.0301^*$ 、中級前半:  $F(2,137)=23.063$ ,  $p=.0000^{****}$ 、中級後半:  $F(2,137)=8.443$ ,  $p=.0003^{****}$ )。

ライアン法 (5%水準) による多重比較を行ったところ、以下のような結果になった。

初級前半: 上位群 = 中位群 < 下位群

初級後半: 上位群 < 中位群

中級前半: 上位群 > 中位群 > 下位群

中級後半: 上位群 > 中位群 = 下位群

ここからは、全体的には初級語で 7 割を占めること、下位群は初級前半語の割合、中位群は初級後半語の割合が他と比べて高いこと、中級前半語はレベルが上がるにつれて使用割合が高くなること、上位群になると中級後半語の割合が上がるのがわかる。

#### (4) 考察

上記結果を踏まえ、1 で示した課題について考察する。

課題 1 「日本語能力と作文の自動評価結果は関係があるのか。日本語能力と文章能力には関係があるのか。」については、日本語テストは全ての自動評価の結果と相関が見られ、自動評価は日本語能力の影響が強いということがわかった。自動評価の中の「目的・内容」や「構成・結束性」といった文章能力も日本語テストとの相関が見られたことから、文章能力はある程度の日本語能力がなければ発揮できないという可能性が示唆された。L2 と第一言語とでは言語処理に要する負荷が異なると考えられるが、佐久間 (2008) は、言語処理におけるワーキングメモリーについて母語の優位性を指摘し、L2 学習者も習熟度が上がれば母語話者に近づくことを示している。田中・久保田 (2014) は、ワーキングメモリーという点から考えると、言語能力が高ければ文章を書く際に文章能力を十分発揮できるが、言語能力が十分でなければ発揮できないとしており、本研究の結果はこれを支持するものとなった。

課題 2 「各自動評価は互いに関係があるのか。」については、Goodwriting Rater のホリスティック評価と jReadability、jWriter による評価はそれぞれ相関があったことから、これらの自動評価サイトは類似の評価結果を示すと考えられる。自動評価はまだ完全なものとは言えないため、場合によっては、ある自動評価サイトの出す結果に問題がある可能性もある。そのような場合、複数の自動評価サイトを使って検証することで、不完全性を互いに補うこともできるのではないだろうか。

課題 3 「作文の総合的評価とテキストデータにはどのような関係があるのか。」については、作文の総合的評価が高いものは、作文全体の長さや一文の長さが長くなる一方で文数は減り、形態素の数が多くなるということがわかった。これは田中他 (2017) の、上位群はエッセイ全体が長く、一文も長いこと、上の群ほど異なり形態素数も総形態素数も多くなることという結果、および、長谷川 (2004) の、作文評価は延べ語数・異なり語数と高い相関があるという結果と一致する。一方、漢字率との関係は認められなかった。この結果からは、作文を書かせる際には、①作文全体を長くする、②長い文を書く意識を持つ、③多様な語彙を使用する、という指導をすることが効果的だということが言えるだろう。

課題 4 「日本語能力とテキストデータにはどのような

関係があるのか。」については、「日本語能力」も課題 3 の「作文の総合的評価」と同様の結果となった。今回の調査においては、漢字の使用は、作文の総合的評価とも日本語能力とも相関関係が認められなかった。これは協力者の 72% が漢字圏の留学生で、下位群も漢字使用率が低くはなかったことが影響しているものと思われる。

日本語能力によって、上位群・中位群・下位群に分けて詳細に分析したところ、上位群は下位群より作文全体が長く、下位群は短文の羅列で、日本語力が上がるほど長い文が増えること、日本語力が上がるにつれて語彙も豊富になることがわかった。日本語力が上がるにつれて総形態素数が多くなることと、漢字率は日本語力による差がないことを合わせて考えると、下位群は同じ漢字語彙を繰り返し使用しているものと思われる。日本語力の低い学習者には、同じ漢字語彙を何度も使わず、多様性を持たせるよう意識させるべきであろう。

語彙レベル構成の結果からみると、下位群は初級前半語、中位群は初級後半語の割合が他と比べて高く、中級前半語はレベルが上がるにつれて使用割合が高くなって、上位群になると中級後半語の割合が上がるということがわかった。下位群・中位群は自分が知っている簡単な日本語だけでなく、意識してまずは中級前半語彙を使用すること、上位群も中級後半語彙の使用をさらに増やすよう指導することが必要であろう。

3 つの群に分けた分析で上位群と中位群の差が見られたのは語彙の豊かさのみであった。一文の長さに関しては、日本語母語話者と比べると、上位群でもかなり短いことから、長く複雑な文の生成ができるまでには至っていないことがわかった。つまり、中位群・上位群に対しては、語彙力の向上に目を向けるだけでなく、文の構造においてもさらに指導する必要があるだろう。

## 5. まとめと今後の課題

本稿では、大学における留学生の日本語プレースメントテストの結果と、Goodwriting Rater、jReadability、jWriter という日本語ライティングの自動評価サイトを利用し、日本語能力、作文の自動評価結果、テキストデータの間関係を分析し、その結果から作文指導に何が必要であるかを考察した。

今回使用した自動評価サイトは、評価だけでなく、テ

キストデータも瞬時に知ることができ、作文指導の際には有効な情報となると思われる。IT 技術が日々進歩する現代において、それらをいかに活用するかを模索することは意味があるだろう。

今回は量的分析に留まったが、今後は各レベルの作文の質的分析も行い、文章能力についても調査を進めたい。

### 【補注】

- 1) Goodwriting Rater は、評価結果の精度に関しては 100%保証されるものではないとしており、jWriter では「初級の作文に対しては精度が低い」(李他 2017) とされている。
- 2) SPOT とは Simple Performance-Oriented Test のことで、日本語力判定に広く使われている。
- 3) Goodwriting Rater では、1-2・3・4・5-6 の 4 段階で評価を出すため、本調査では下から 1、2、3、4 点として分析した。また、jReadability では文章難易度を初級前半～上級後半の 6 段階で評価しているため、下から 1～6 点とした。同様に、jWriter の入門、初級、中級、上級、超級は 1～5 点とした。

本研究は、JSPS 科研費 26284074「日本語ライティング評価の支援ツール開発：「人間」と「機械」による評価の統合的活用」(田中真理代表) の助成を受けたものである。

### 【引用・参考文献】

- 宇佐美洋・田中真理・徳井厚子 (2012)「評価の『個人差』に着目することの意味—より深い自己認識につなげるための評価論—」『ヨーロッパ日本語教育』16、36-50
- 佐久間康之 (2008)「ワーキングメモリ内の英語処理に関する母語話者と日本人外国語学習者との比較研究」『日本言語テスト学会研究紀要』11、173-188
- 田中真理・阿部新・影山陽子・佐々木藍子・坪根由香里 (2017)「ヨーロッパ日本語学習者のライティング(エッセイ)分析：総合的評価とマルチプルトレイト評価結果を参照して」『ヨーロッパ日本語教育』22、ヨーロッパ日本語教師会、75-92
- 田中真理・久保田佐由利 (2014)「アカデミック・ライティングの構成面について：L1, L2 双方向からの考察」『2014 CAJLE Annual Conference Proceedings』163-173
- 田中真理・坪根由香里 (2011)「第二言語としての日本語小論文における good writing 評価—そのプロセスと決定要因—」『社会言語科学』14(1)、210-222
- 田中真理・初鹿野阿れ・坪根由香里 (1998)「第二言語としての日本語における作文評価—『いい』作文の決定要因—」『日本語教育』99、60-71
- 坪根由香里・田中真理 (2015)「第二言語としての日本語小論文

評価における『いい内容』『いい構成』を探る—評価観の共通点・相違点から—『社会言語科学』18(1)、111-127

長谷川守寿 (2004) 「プレースメントテストにおける作文の評価と形態的特徴の関係について」『筑波大学留学生センター日本語教育論集』19、1-20

原田朋子 (2019) 「日本語母語話者と上級日本語学習者の小論文の比較：テキストマイニング手法と目視による分析を通して」『同志社大学日本語・日本文化研究』16、1-15

李在鎬 (2016) 「日本語教育のための文章難易度に関する研究」『早稲田日本語教育学』21、1-16

李在鎬・長谷部陽一郎・迫田久美子 (2017) 「人工知能の仕組みを利用した学習者作文評価システム『jWriter』—I-JAS を利用した試み」『2017 年度日本語教育学会秋季大会予稿集』289-291

<自動評価サイト>

GoodWriting.jp 日本語ライティングの自動評価システム

<https://goodwriting.jp/wp/>

jReadability 日本語文章難易度判別システム

<https://jreadability.net/sys/ja>

jWriter 学習者作文評価システム

<https://jreadability.net/jwriter/>